DISCUSSION

J.N.K. Rao, Carleton University, Ottawa, Canada

Section I of the paper by Grant Capps gives some theory for a generalized unequal probability sampling design which includes the usual with and without replacement designs as special cases. An interesting application to the Current Population Survey is given in Section II. The remaining two sections (III and IV) investigate a sample selection method which is a compromise between the one unit per stratum and the two units per stratum designs.

The generalized estimator of the population total Y considered in Section I is given by

$$\hat{Y} = \sum_{i=1}^{N} \frac{t_i}{E(t_i)} y_i \,, \tag{1}$$

where $t_i (i = 1,\ldots,N)$ is the number of times the i-th population unit is included in a sample of fixed size $n\ (\Sigma t_i = n)$. Capps derived the variance of $\hat{Y}$ and two unbiased variance estimators from first principles. In this connection, it may be of interest to note that these results can be obtained simply from a general theorem (Rao and Vijayan [2]) which, in addition, gives the necessary form of nonnegative unbiased estimators of MSE. A general linear estimator of Y is given by

$$\hat{Y} = \sum_{i=1}^{N} d_{is} y_i \tag{2}$$

where s denotes a sample selected according to a design $p(s)$, and the weights d_{is} in (2) are such that $d_{is} = 0$ if $i \notin s$. We have the following general theorem:

Theorem. Suppose the mean square of $\hat{Y}_d$ becomes zero when the ratios y_i/w_i are all equal for some constants $w_i (\neq 0)$. Then

(a) $MSE(\hat{Y}_d)$ reduces to

$$MSE(\hat{Y}_d) = -\sum_{i<j}^{N}\sum d_{ij} w_i w_j (z_i - z_j)^2 \tag{3}$$

where $z_i = y_i/w_i$ and

$$\begin{aligned} d_{ij} &= E(d_{is} - 1)(d_{js} - 1) \\ &= \Sigma_s p(s)(d_{is} - 1)(d_{js} - 1)\,; \end{aligned} \tag{4}$$

(b) a nonnegative quadratic unbiased estimator of $MSE(\hat{Y}_d)$ is necessarily of the form

$$mse(\hat{Y}_d) = -\sum_{i<j}\sum d_{ij}(s) w_i w_j (z_i - z_j)^2 \tag{5}$$

where $d_{ij}(s) = 0$ if s does not contain both units i and j, and

$$E(d_{ij}(s)) = \sum_{s \ni i,j} p(s) d_{ij}(s) = d_{ij}\,, \quad i < j\,. \tag{6}$$

Equation (6) is the unbiasedness condition, and selected choices of $d_{ij}(s)$ satisfying (6) lead to unbiased estimators of $MSE(\hat{Y}_d)$. If $\hat{Y}_d$ is unbiased for Y as in the case of $\hat{Y}$, then $E(d_{is}) = 1$ and (4) reduces to

$$d_{ij} = E(d_{is} d_{js}) - 1\,. \tag{7}$$

We now illustrate the application of (3) - (7) to the estimator $\hat{Y}$ given by (1). The condition of our Theorem is satisfied with $w_i = E(t_i)$ and $y_i/w_i = c (\neq 0)$, since $\hat{Y}$ reduces to $c\,\Sigma t_i = cn$, a constant. Noting that $d_{is} = t_i/E(t_i)$ for $\hat{Y}$, we get from (4)

$$d_{ij} = cov(t_i\,, t_j)/\{E(t_i)E(t_j)\}\,, \tag{8}$$

and (3) reduces to the formula (5) of Capps:

$$V(\hat{Y}) = -\sum_{i<j}\sum cov(t_i\,, t_j)(z_i - z_j)^2 \,. \tag{9}$$

The choice

$$d_{ij}(s) = d_{ij} \frac{t_i t_j}{E(t_i t_j)} \tag{10}$$

satisfies (6), and (5) reduces to

$$v(\hat{Y}) = -\sum_{i<j}\sum \frac{t_i t_j}{E(t_i t_j)} cov(t_i\,, t_j)(z_i - z_j)^2 \tag{11}$$

which agrees with the formula (7) of Capps. The variance estimator (6) of Capps does not belong to the necessary class of nonnegative unbiased variance estimators, viz. (5).

The compromise scheme in Section III was obtained by choosing Scheme I (one unit per stratum design) with probability p and Scheme 2 (Durbin's scheme) with probability $1 - p$ $(0 \leq p \leq 1)$ and then selecting a sample of $n = 2$ units according to the chosen scheme. The variance formulae derived in Section III E (for the unconditional estimator $\hat{Y}_p$) and in Section III F (for the conditional estimator $\hat{Y}_c$) can be obtained simply from the general formulae (3) and (5) with the choice $d_{ij}(s) = d_{ij}/\pi_{ij}(p)$, $i < j \in s$. It also follows that (28) and (34) are the only possible nonnegative unbiased variance estimators for $\hat{Y}_p$ and $\hat{Y}_c$ respectively.

Fuller [1] has also proposed the compromise scheme, but confined himself to simple random sampling designs in which case $\hat{Y}_p$ and $\hat{Y}_c$ both reduce to $N\bar{y}$, where $\bar{y}$ is the sample mean. Fuller proposed an alternative method which appears preferable to the compromise scheme. The method is approximately as efficient as the one unit per stratum design and yet provides unbiased variance estimators. An extension of the alternative method to unequal probability sampling was also given.

REFERENCES

[1] Fuller, W.A. (1970). "Sampling with random stratum boundaries", Journal of the Royal Statistical Society, Ser. B, 32, 209-26.

[2] Rao, J.N.K. and Vijayan, K. (1977). "On estimating the variance in sampling with probability proportional to aggregate size", Journal of the American Statistical Association, 72, 579-84.

The paper by Isaki and Pinciaro gives useful empirical results on the relative performances of seven variance estimators for PPS systematic sampling. However, the study was confined to just one population, viz. mobile home dealers canvassed in the 1972 Census of Retail Trade. It would be useful if the study is extended to cover other real populations. Model-based investigations would also throw further light on the properties of the variance estimators. A model-based variance estimator proposed by Hartley [1] is not included in the study.

REFERENCES

[1] Hartley, H.O. (1966). "Systematic sampling with unequal probability and without replacement", Journal of the American Statistical Association, 61, 739-48.